Related to other papers in this special issue	23 (p230); 15 (p151); 13 (p131); 21 (p208)
Addressing FAIR principles	F, A, I, R

FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources

Hana Pergl Sustkova[1†], Kristina Maria Hettne[2], Peter Wittenburg[3], Annika Jacobsen[4], Tobias Kuhn[5], Robert Pergl[6], Jan Slifka[6], Peter McQuilton[7], Barbara Magagna[8], Susanna-Assunta Sansone[7], Markus Stocker[9], Melanie Imming[10], Larry Lannom[11], Mark Musen[12] & Erik Schultes[1]

[1]GO FAIR International Support and Coordination Office, Leiden, The Netherlands

[2]Centre for Digital Scholarship, Leiden University Libraries, Leiden, The Netherlands

[3]Max Planck Computing and Data Facility, Gießenbachstraße 2, 85748 Garching, Germany

[4]Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands

[5]Department of Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 11051081 HV Amsterdam, The Netherlands

[6]Czech Technical University in Prague, Faculty of Information Technology (FIT CTU), 160 00 Prague 6, Czech Republic

[7]Oxford e-Research Centre, Department of Engineering Sciences, University of Oxford, Oxford OX13PJ, UK

[8]Environment Agency Austria, A-1090 Vienna, Austria

[9]TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

[10]SURF, Utrecht 3511 EP, The Netherlands

[11]Corporation for National Research Initiatives (CNRI), Reston, Virginia 20191, USA

[12]Stanford Center for Biomedical Informatics Research, Stanford, CA 94305, USA

† Corresponding author: Hana Pergl Sustkova (E-mail: hana.pergl@go-fair.org, ORCID: 0000-0002-4462-6465).

chinaXiv:202211.00178v1

ABSTRACT

The FAIR principles articulate the behaviors expected from digital artifacts that are Findable, Accessible, Interoperable and Reusable by machines and by people. Although by now widely accepted, the FAIR Principles by design do not explicitly consider actual implementation choices enabling FAIR behaviors. As different communities have their own, often well-established implementation preferences and priorities for data reuse, coordinating a broadly accepted, widely used FAIR implementation approach remains a global challenge. In an effort to accelerate broad community convergence on FAIR implementation options, the GO FAIR community has launched the development of the FAIR Convergence Matrix. The Matrix is a platform that compiles for any community of practice, an inventory of their self-declared FAIR implementation choices and challenges. The Convergence Matrix is itself a FAIR resource, openly available, and encourages voluntary participation by any self-identified community of practice (not only the GO FAIR Implementation Networks). Based on patterns of use and reuse of existing resources, the Convergence Matrix supports the transparent derivation of strategies that optimally coordinate convergence on standards and technologies in the emerging Internet of FAIR Data and Services.

The nice thing about standards is that there are so many of them to choose from. – Grace Hopper[1]

1. INTRODUCTION

Although the FAIR Principles were published in 2016 [1], broad community-wide discussions around FAIR implementations are only now beginning to emerge. However, a plethora of technologies and standards have been in development over decades that have aimed towards or support, in one way or another, the FAIR behavior of digital artifacts. For example, the European Strategy Forum on Research Infrastructures (ESFRI) has supported multilateral, interlocking initiatives leading to research infrastructures with an array of well-used technical solutions for preserving, locating and reusing research data [2]. The Research Data Alliance (RDA), launched in 2013 and now with 87 active Working/Interest Groups, has continuously supported active development of data sharing and reuse (see specifically a recent overview provided by the FAIRsFAIR Horizon 2020 project, indicating the RDA groups having especially strong relevance for FAIR[2]). FAIRsharing alone lists over a thousand FAIR-related data and metadata standards, minimal information checklists, terminologies, models, schemas and formats that are already in use [3,4,5]. Furthermore, over several decades the W3C and especially working groups of its former Semantic Web Activity [6] and current Data Activity [7], have developed technologies such as the Web Ontology Language [8] which implement some of the FAIR principles (e.g., principle I1 makes explicit reference to languages for knowledge representation). Hence, a community's commitment to FAIR implementation is much less a question of setting out to create novel technologies, than it is to achieve optimal reuse of the manifold existing solutions.

---

[1] Quoted in Metadata by Jeffery Pomerantz, The MIT Press Essential Knowledge Series (2015).

[2] https://docs.google.com/spreadsheets/d/1QmWXm4Wil0bbMU8WYrnHFe4gbSzT-CDpbATUukESgEM/edit?usp=sharing.

By "optimal reuse" we mean a pattern of adoption of implementation solutions that minimizes costly "re-invention of the wheel" while at the same time maximizes data interoperation. A key factor in optimal reuse is the identification of a "critical mass" of users of a given technology such that they are likely to attract still other communities to reuse the same approach. In the limit of this process, numerous FAIR implementation choices could evolve into broadly accepted default standards and make it increasingly straightforward for anyone to then create and exchange FAIR digital artifacts[③]. Default standards, if they could be encouraged to emerge, would create opportunities for service providers (be they public or private) to invest in scalable solutions.

In the interest of maximizing reuse, the GO FAIR Initiative has been working with diverse communities to (1) compile a comprehensive inventory of existing resources which communities have explicitly chosen to implement FAIR and (2) analyze these resources to identify optimal reuse strategies. The result is an open and FAIR collection, or Matrix, of communities and the FAIR-related resources they use. The Matrix permits a transparent, coordinated and bottom-up community convergence [9] to agreed-upon implementations that when taken together, collectively compose a broad Internet of FAIR Data and Services.

Originally inspired by a simple questionnaire [10] and spreadsheet [11] exercise at the annual GO FAIR Implementation Network meeting (Leiden, January 2019, with about 100 data experts present) [12], a small coalition of FAIR developers[⑧] came together (June 2019) to build a professional, robust FAIR Convergence Matrix [13]. The Convergence Matrix is essentially a dynamic table of communities of practice (the "columns" in the Matrix) composed of self-identified entities such as GO FAIR Implementation Networks (INs), funded research projects, consortia, alliances, companies, etc. and the resources (the "rows" in the Matrix) they have chosen to implement FAIR data and services. These resources include policies, technologies, data, metadata, ontologies, data models, and standards. The Matrix is designed to prompt, capture and share the implementation choices explicitly declared by each community. The resulting Matrix content will be in machine-actionable format and will be openly reported, so that usage statistics of FAIR-related resources can be assessed for any and all stakeholders who wish to voluntarily declare their own strategies for "going FAIR". Owing to the dynamic and rapidly evolving landscape of standards and technologies, the Matrix will be accessible in real-time, and will trigger alerts that are appropriate for different stakeholders. As it develops, the FAIR Convergence Matrix will provide an informative global overview of the community practices that are in use to implement FAIR data and services. Overall, the Matrix will:

---

③ By analogy, TCP/IP are Internet protocols that became so universally accepted that they have "faded into the background" such that the vast majority of users today are blissfully ignorant of their existence. Yet every "click" on a link, Google search, Tweet, and email are executed on a series of networks all interconnected by the protocols of TCP/IP.

⑧ The Convergence Matrix Coalition was formally launched on June 13 and has now regular bi-monthly meetings. The coalition includes representatives of the GO FAIR International Support and Coordination Office, the Centre for Digital Scholarship at Leiden University Libraries, the Free University Amsterdam, Czech Technical University in Prague, FAIRsharing (under the GO FAIR StRePo IN, and as the Research Data Alliance (RDA) FAIRsharing WG), GEDE (a collaboration of 47 research infrastructures (RIs), mostly ESFRI and ERIC initiatives), the CS3 consortium, ENVRI-FAIR (a cluster project in the environmental sciences with 20 RIs involved) and the RDA. Ongoing efforts recorded at https://osf.io/xqfb9/.

- Gather existing solutions towards FAIR and encourage a multiplicity of stakeholders to formulate and coordinate implementation choices.
- Foster cross-disciplinary interoperability via the reuse of existing resources. Good practices applied in one specific scientific domain or geographical region can be considered for adoption by other communities of practice. This kind of alignment should preempt rampant and wasteful "reinvention of the wheel" and prompt, where necessary, careful consideration (and justification) for the development of new approaches.
- Respect the geographical and disciplinary boundaries of the community, while encouraging interoperation convergence toward globally-accepted frameworks.
- Provide frameworks for organizing an otherwise rugged landscape of complex and interrelated standards, technologies, data and services. For example, elements of the Matrix could be ordered by the FAIR principles, Maturity Indicators or functions that are generic (e.g., identifier resolution services) *versus* domain-specific (e.g., provenance metadata components describing the methodological approach to sampling human subjects in language research).
- Provide the raw data upon which recommendations for FAIR best practices can be systematically and transparently developed. For example, the European Open Science Cloud (EOSC) FAIR working group [14] could use the Convergence Matrix to assess the current usage of existing Persistent Identifier services to better formulate its own recommendations for the EOSC.
- Assist communities of practice (ordinarily via local Data Stewardship Competency Centers®) when formulating FAIR data stewardship plans.
- Deliver a comprehensive inventory of standards needed when formulating FAIR maturity indicators and evaluation/certification systems to assess the levels of FAIRness among digital resources.

2. DEVELOPMENT OF THE FAIR CONVERGENCE MATRIX: QUESTIONNAIRE AND WIZARD INTERFACE

The data captured by the Convergence Matrix will be exposed as an open and FAIR knowledge graph that captures how resources are currently being used by communities that implement FAIR Data and Services. This means that the knowledge graph can be mirrored in multiple instances by able and willing hosts, and that Matrix inputs and outputs can be created in any environment that produces the relevant FAIR data. Furthermore, any third-party service will be welcomed to derive from the knowledge graph useful information serving the FAIR-related needs of its own stakeholders. This being said, the Convergence Matrix coalition is presently leveraging existing tools and methods that are well established among its members to build a practical, user-friendly Convergence Matrix. We highlight the approach and some key features below and in Figure 1.

® A term currently used by the preparatory DSCC GO FAIR Implementation Network denoting any institutional competence center dealing with FAIR Data Stewardship issues.
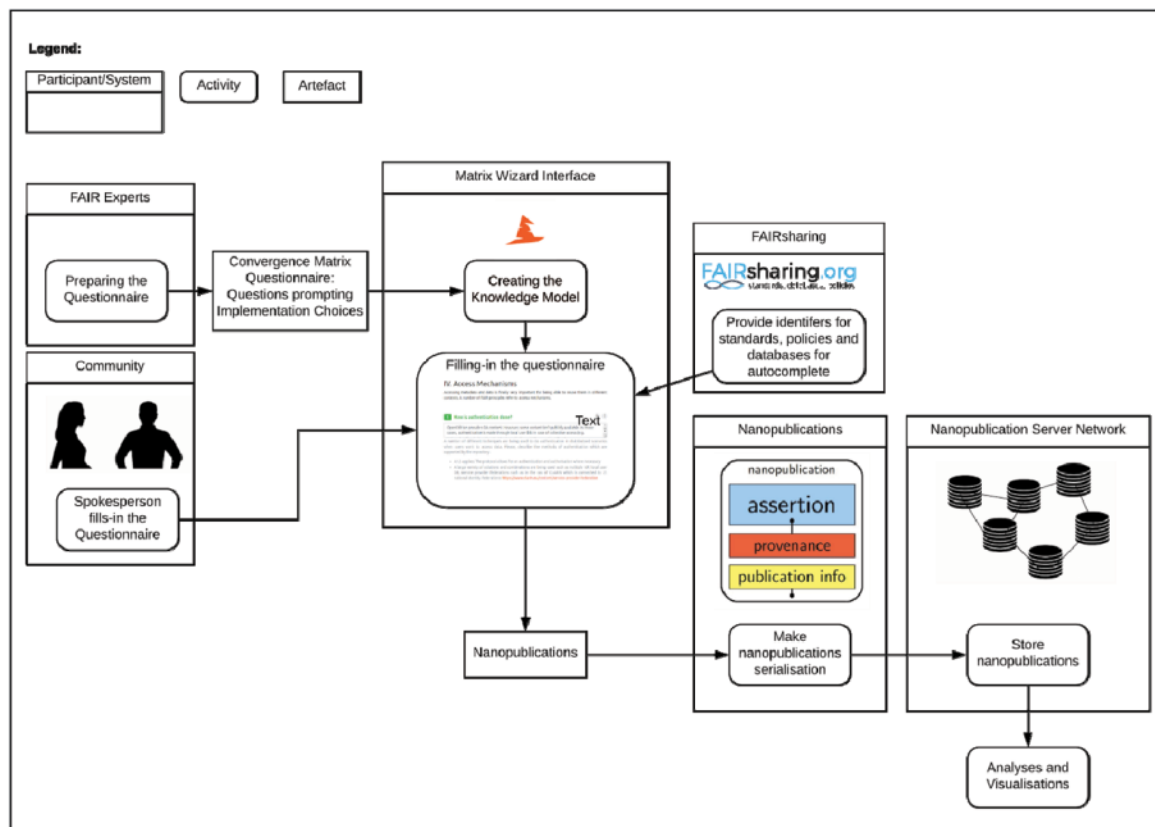
Figure 1. Convergence Matrix Process Overview. The questionnaire is composed and maintained by FAIR Experts, an effort was made to ensure broad coverage of technologies and other Resources and how they relate to each of the FAIR principles. The questionnaire is encoded in a machine-readable Wizard Knowledge Model, which then exposes the questions in a user-friendly interface (screenshot). The community spokesperson registers in the Wizard, completes a few questions profiling the community, then begins to answer the 61 questions in the questionnaire. Default answers, drop-downs and autocomplete make the completion of the form easier and help achieve the machine readability. At some point in the future, Communities and trusted third-parties (e.g., funding agencies, publishers, data stewards, etc.) could publish customized Knowledge Models that will offer recommendations on, or even require the use of, certain Resources. This function could be a powerful driver of convergence. The drop-down and autocorrect is provided by FAIRsharing. The data input by the Community Spokesperson is captured as stand-alone nanopublications (capturing an assertion about the "Implementation Choice Made" and documenting the decision with a collection provenance metadata). The nanopublications will be made available on the distributed nanopublication server network, and will be available to any other organizations for hosting and serving. The resulting open knowledge graph is generated from the stored data and can be viewed as a public good, advising a myriad of decisions needed to launch and sustain the Internet of FAIR Data and Services.

In principle, widespread convergence onto standards and data for FAIR can be accomplished by asking communities of practice to simply list all of the digital resources they use (or are building) relevant to the FAIR principles. With such an inventory, communities can then collectively choose to reuse the FAIR-related resources that are already widely used or that demonstrate some superior attributes. For example: The *GO FAIR Chemistry Implementation Network* [15] lists the IUPAC International Chemical Identifier system for encoding molecular information (addressing FAIR Principles F1); The *Metrology community* lists developing metadata standards for electron microscope image data (addressing FAIR Principles F2 and R1.2); the *Rare Disease community* [16] lists the Orphanet Rare Disease Ontology and the Human Phenotype Ontology in linking disease registries (addressing FAIR Principles I1 and R1.3); still other communities can list their choice for access protocols (addressing FAIR Principles A1) or machine-actionable licensing (addressing FAIR principles R1.1).

In practice however, it is very helpful if these diverse communities are prompted to supply answers to a standard set of questions that systematically and comprehensively cover the FAIR principles. By standardizing such questions, asking respondents to carefully consider their options and then make explicit decisions in each case, it is easier to spot patterns of use and reuse that can later be exploited for the purposes of widespread convergence on standards and technology.

Starting with the initial and somewhat embryonic GO FAIR questionnaire, the Convergence Matrix coalition soon developed advanced questionnaires that were more rigorous and comprehensive and were carefully embedded in terminology familiar to domain-experts, IT professionals and data stewards [17]. In this process, the questionnaire has evolved into a lengthy and complex document, prompting implementation decisions that are both generic and highly specific to particular research domains. As such, completing the questionnaire is at best tedious and at worst fraught with uncertainty and potential ambiguity and errors. Furthermore, FAIR implementations will themselves never be static, but are undergoing continual development requiring regular updates to both the questionnaire and the responses offered by communities. In order to address the questionnaire complexity and to make it easier for respondents to complete and maintain community profiles, the questionnaire has been represented using a dynamic, Web interface originally developed as part of the Data Stewardship Wizard [18,19].

The Wizard questionnaire engine provides the possibility to capture the questions and the answers using semantically-enabled drop-down menus and auto-complete functions. This feature helps the user to navigate the decision-making process and efficiently record, unambiguously, the answers to potentially hundreds of questions. The drop-down/auto-complete values are taken from FAIRsharing, which provides globally unique and persistent identifiers and metadata descriptions of FAIR-related standards, repositories and data policies. The combination of FAIRsharing and the Wizard allows communities to select from, and reuse existing resources in a managed and machine actionable way. The Wizard interface also provides links to explanatory and reference information that can guide decision making.

The questionnaire structure is captured in the so-called *Knowledge Model* of the Wizard. The Knowledge Model has a hierarchical structure, so questions can be dynamically presented as needed, based on previous

responses. The Convergence Matrix questionnaire Knowledge Model is itself machine readable (exportable in JSON format), can be easily edited and subsequent versions of the questionnaires migrated and tracked as a FAIR resource. The answers are stored in a document database and are transformed into JSON-LD for an interoperable RDF representation suitable for further analysis and knowledge engineering. Beginning in July 2019, the Matrix coalition started testing the advanced questionnaires and Wizard interface in real-world cases that involve more than 30 European Research Infrastructures including those participating in the ENVRI-FAIR project [20]. Following testing of the questionnaire and its refinement the Convergence Matrix questionnaire Knowledge Model can be registered in FAIRsharing.

3. DEVELOPMENT OF THE FAIR CONVERGENCE MATRIX: SEMANTIC DATA MODELS

In a concerted effort to ensure the Convergence Matrix would itself be openly accessible as a FAIR resource, the terminology of the questionnaire is in the process of being aligned with that of the GO FAIR Reference Ontology© (GFRO) [21] to maintain conceptual precision for both humans and machines. The GFRO provides a means of ontological clarification of definitions and relations regarding key development trends in FAIR, including the definitions and relations of the FAIR principles themselves, FAIR Digital Artifacts, Maturity Indicators for evaluating the level of FAIRness in any digital resource, FAIR Certification and the FAIR Convergence Matrix. The GFRO represents a shared understanding primarily targeted for humans, but also provides a reference point for making and synchronizing technical ontologies such as semantic data models [22] and database schemas. Our approach is inspired by leading efforts in Applied Ontology as described e.g., in [23].

These formal data elements are captured using nanopublications [24,25], small packages of information in a formal representation language (RDF) that come with structured provenance and metadata. Methods and infrastructures have been created that allow us to reliably identify, version, publish, and retrieve these small data snippets, as well as to combine them to larger data sets [26] rendering the data in the Convergence Matrix to a large degree, FAIR. The nanopublication data model is based on the "FAIR Vocabulary" [27]. As independently emerging conceptual models in the FAIR space, the GFRO and the FAIR Vocabulary currently exhibit some important inconsistencies in the concept names they use and how they model relations. Resolution of the foundational ontology with the semantic data model promises great benefits to reusability however, and is thus a key focal area in Convergence Matrix development. Box 1 briefly describes the GFRO and demonstrates how the Convergence Matrix may be used to drive convergence.

The nanopublications are used throughout to capture and track all data elements of the Convergence Matrix in a machine-actionable manner as linked data. This includes the FAIR principles themselves, the questions in the questionnaire, community profiles (including the Spokespersons), the resources, and the commitments supplied by the user (declared choices and challenges). As linked data, all these data elements in the Convergence Matrix have globally unique, resolvable and persistent identifiers and associated

---

© Registered in FAIRsharing as https://fairsharing.org/bsg-s001394/.

metadata provided by FAIRsharing. The unambiguous and persistent identification of resources is a key feature of the Convergence Matrix, making implementation choices easier to define and reuse.

Nanopublications allow the collected linked data of the Convergence Matrix to be exposed in multiple FAIR endpoints for open, real-time, machine-accessible data sharing. By publishing these data on the decentralized nanopublication server network[⑦], they are made Findable (F) and Accessible (A) through the network and its API [28]. The nanopublication linked data and representation allows for interoperability (I) and detailed formal provenance ensures Reusability (R). Together with the methods and infrastructures mentioned above, nanopublications now help us to publish this model and its instantiations in a FAIR manner.

Box 1 Using the Matrix to Catalyze Convergence on FAIR Implementations:

We use bold font to indicate concepts defined in the Reference Ontology of the Matrix.

The Convergence Matrix is composed of two fundamental data elements: Communities of Practice, or *Communities* and the FAIR-Enabling Digital Resources, or *Resources*.

A Resource provides a function needed to achieve FAIRness and is explicitly linked to one or more FAIR principles. Examples of Resources include data and metadata standards (such as vocabularies, formats), data repositories, policies, and training materials. Communities have been intentionally defined in a permissive manner in the Convergence Matrix to encourage as broad participation as possible: they are self-identified entities (not limited to, for example, formal legal entities); they can be large or small (although as a consensus-building exercise, communities must be composed of more than one person); they are fluid (in that communities can grow, shrink, merge and split); and they may have multiple identity principles (e.g., by topical domain, by project organizations, by institutional affiliations). Every community has a **Spokesperson** (domain-specific FAIR data steward or coordinator) who takes responsibility for completing and updating the questionnaire on behalf of the Community, and who can be contacted by others when there are technical questions about Resource use.

When a Community wishes to implement the **FAIR principles**, it must first make careful **Considerations** about a number of generic, domain-specific, and cross-domain **Requirements** and **Constraints**. These **Considerations** clarify the FAIR-related needs of the Community and translate into the many Implementation **Choices to be Made** regarding Resources that enable maximum reuse and interoperation.

The Convergence Matrix workflow begins when a Community is presented, via the **Questionnaire**, with a list of Choices to be Made. In each instance, the **Choice Made** by the Community Spokesperson is for a Resource that is either already **Available for Use** or if the Resource does not yet exist, then the

---

⑦ The server network could be managed by other willing and trusted organizations such as the Centre for Digital Scholarship at Leiden University Libraries or the Dutch National Computer Center.

Community may announce a **Commitment to Develop** it. The Commitment to Develop can be self-directed back to the Community, or "given away" to other members of the stakeholder community who are able and willing to launch new Resource development projects. Hence, for any given Implementation Choice to be Made presented by the Questionnaire via the Wizard interface, a Community must either make a **Commitment to Reuse** a Resource that is Available for Use or make **a Challenge Proposal** to build an entirely new Resource.

In this conceptualization, all existing Resources that are now Available for Use, were at one time proposed as a Challenge to Develop. In the Convergence Matrix, these Community relations to Resources are explicit and publicly registered as Implementation Choices Made. As such, the Convergence Matrix encourages wherever possible, reuse of existing resources but will also track the accumulated list of implementation challenges, providing a FAIR development "road map" for the community-at-large.

As an illustration of how this conceptual model can engender convergence in FAIR standards and technologies, consider three Communities: Agriculture (Community A), Biology (Community B) and Chemistry (Community C). Each Community has developed Resources relevant to their knowledge domains: Community A has developed its own authoritative soil ontology (as part of the Global Soil Partnership [29]); Community B has developed multiple standards for the identification of biological species, including the Encyclopaedia of Life [30] and GBIF [31]; Community C has developed its own chemical identifier schema called the International Chemical Identifier [32]. Naturally, upon entering the Convergence Matrix, each Community declares its commitment to reuse its own domain-relevant Resources. In doing so, they make these Resources visible to all other Communities in the Matrix, and the Questionnaire can be automatically extended to then prompt other Communities to also Consider uptake. Going forward, each Community will be alerted by the growing Questionnaire to make choices regarding the reuse of Resources from each of the other Communities. For example, Community A will be prompted by the Questionnaire to choose identification schema for biological species (created by the trusted authorities in Community B) and chemicals (likewise, created by trusted authorities in Community C). This informed, cross-disciplinary decision making will be prompted by the Matrix Questionnaire even when the Resources may not seem to have direct or immediate relevance. For example, Communities B and C will also be prompted by the Questionnaire to also choose a soil ontology (created by the trusted authorities in Community A). The cross-disciplinary reuse of these resources among the communities (rather than their continuous redevelopment) would constitute a tremendous cost savings and convergence accelerator. Furthermore, Communities will be systematically prompted to commit to the use of more generic Resources, such as schemata for access protocols, data licensing, geo-location, time/date records, citation of research articles, or the identification of people (using ORCID for example). Hence, given community-specific considerations for data interoperation and reuse, prompted by a systematic and comprehensive Questionnaire, and guided by the collective implementation choices and challenges of other Communities, Communities A, B, and C will have the knowledge required to optimally choose, test and revise its commitments to FAIR implementations options.

4. FAIR CONVERGENCE MATRIX ANALYTICS

Given the machine-actionable representations for communities, resources, choices, and challenges, it is possible to query from many different points of view, and filter/sort this information in a variety of ways. For example, the knowledge graph of the Convergence Matrix can be parsed by community (topical or knowledge domain, size, geographical location), by resource (technologies, training / educational materials, policies), or by the FAIR principles themselves. Attributes such as the openness of resources, and indicators of resource maintenance, stability and sustainability as well as usage trends can all be analyzed. These, and many other possible "views" on the Convergence Matrix will make it possible to run sophisticated analyses on the resulting Matrix data, reflecting common usage patterns among topical domains, geographic regions, larger and smaller communities of practice. With time, Matrix analytics will also demonstrate the evolving character of FAIR implementations, revealing trends in usage as new technologies emerge, potentially replacing existing methods. In turn, these statistical patterns can be used to inform national and international policy makers and funding priorities for optimizing the reuse of existing resources for implementing FAIR data and services. Such analyses will initially be performed by the coalition partners FAIRsharing and the GO FAIR International Support and Coordination Office, but the Matrix output will be openly available and it is anticipated that it will be used also by, for instance, the relevant Working Groups in the EOSC governance (such as FAIR, Architecture, and Rules of Participation), by CODATA or by Data Stewardship Competency Centers at universities.

5. CONVERGENCE MATRIX DEVELOPMENT ROADMAP

The Convergence Matrix coalition has been working in-kind since January 2019 to develop the essential features of the Matrix only briefly covered above. At present, options are being explored to acquire development resources allowing for professionalization and extension of the Matrix, as well as for its deployment accommodating potentially large volumes of community responses. The Convergence Matrix coalition has defined a timeline towards making the Matrix resource broadly available:

Milestone 1: September 2019.

A refined and robust version of the Matrix questionnaire will be available to collect input from the communities of practice using the Wizard technology as an optimized user-interface. The coalition also expects to have collected responses from minimally 10 independent communities representing diverse topical and organizational domains. Responses collected will be analyzed to trigger refinement of the questionnaire. These results are to be presented at CODATA 2019 Beijing.

Milestone 2: October 2019

The Matrix Wizard will feature data capture as nanopublications that will be exposed as an open, FAIR resource. These results are to be presented during RDA's 14th plenary in Helsinki.

Milestone 3: June 2020

Broad launch and open participation of communities of practice. The information will be exposed as a professional (24/7), distributed service.

6. OUTLOOK

We anticipate that international organizations mandated to deal with data policies (such as EOSC-governance bodies and CODATA) will gradually assume responsibility and engender trust regarding the sustainability of the Convergence Matrix as this role is beyond the mandate of GO FAIR. This will enable intellectual driving organizations such as RDA, ESFRI projects, GO FAIR Implementation Networks, International Scientific Unions, Data Stewardship Competency Centers and other domain specific communities of practice to continuously drive the further development of the advanced versions of the Convergence Matrix.

AUTHOR CONTRIBUTIONS

T. Kuhn contributed to the overall conception and the ontology, created the nanopublications, and wrote the text about them. P. McQuilton and S.-A. Sansone worked to connect FAIRsharing with the Wizard, and contributed to the manuscript. K.M. Hettne contributed to the overall conception, created a list of questions for the IN profile and analyzed the results, contributed to the list of questions for the Wizard implementation, contributed to the ontology and nanopublications, and contributed to the manuscript. M. Stocker co-lead the related work in the ENVRI-FAIR project, reviewed and commented on the manuscript. R. Pergl is leading the development team of Wizard and he is the main author and maintainer of the GO FAIR Core Ontology. In this paper, he contributed to the relevant sections and Figure 1. J. Slifka is a member of the Wizard development team and he is the author of the customizations of Wizard for the Matrix project. Together with R. Pergl, he authored the relevant sections. E. Schultes Conception and design of the Convergence Matrix and its early stage implementation; directing the coalition development team. H.P. Sustkova supported the coalition development team; coordinated the team of authors, contributed, reviewed and commented on the manuscript; handled formatting.

ACKNOWLEDGEMENTS

The development of the Wizard technology was funded partially by ELIXIR, the European research infrastructure for life-science data and Institute of Organic Chemistry and Biochemistry AS CR. Considerable amount of work was done in kind by Faculty of Information Technology, Czech Technical University in Prague and by Dutch Techcentre for Life Sciences.

REFERENCES

[1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al. & B. Mons. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3(2016), Article No. 160018. doi:10.1038/sdata.2016.18.

[2] P. Wittenburg, F. de Jong, D. van Uytvanck, M. Cocco, K. Jeffery, M. Lautenschlager, ... & P. Holub. State of FAIRness in ESFRI projects. Data Intelligence 2(2020), 230–237. doi: 10.1162/dint_a_00045.

[3] Sansone, S.A., Susanna-Assunta Sansone, P. McQuilton, P. Rocca-Serra, A. Gonzalez-Beltran, M. Izzo, A.L. Lister, M. Thurston & the FAIRsharing Community. FAIRsharing as a community approach to standards, repositories and policies. Nature Biotechnology 37(2019), 358–367. doi:10.1038/s41587-019-0080-8.

[4] The FAIRsharing registry. Available at: https://fairsharing.org/standards.

[5] P. McQuilton, D. Batista, O. Beyan, R. Granell, S. Coles, M. Izzo, ... & S.-A. Sansone. Helping the consumers and producers of standards, repositories and policies to enable FAIR data. Data Intelligence 2(2020), 151–157. doi: 10.1162/dint_a_00037.

[6] W3C working group on Semantic Web Activity. Information available at: https://www.w3.org/2001/sw/.

[7] W3C working group on Data Activity. Information available at: https://www.w3.org/2013/data/.

[8] Web Ontology Language. Information available at: https://www.w3.org/2001/sw/Specs#OWLSpecs.

[9] P. Wittenburg & G. Strawn (2018). Common Patterns in Revolutionary Infrastructures and Data. Available at: https://www.rd-alliance.org/sites/default/files/Common_Patterns_in_Revolutionising_Infrastructures-final.pdf.

[10] Implementation Network (IN) Profile. Available at: https://docs.google.com/forms/d/e/1FAIpQLScajPBdcMBpVhexvn50CBv9nfQ2p25zWbGL5udcts0gJywJpA/viewform.

[11] GO FAIR IN Profile Matrix. Available at: https://docs.google.com/spreadsheets/d/1MUZn7uh4x5YLPjqxi-V8XubsSEEonQWvx2jBlcyyNdU/edit?usp=sharing.

[12] Report from the GO FAIR Implementation Networks meeting 2019. Available at: https://www.go-fair.org/2019/01/30/go-fair-implementation-networks-meeting/.

[13] A presentation giving an overview of the Matrix. Available at: https://osf.io/ceyvj/.

[14] EOSC Secretariat FAIR working group. Available at: https://www.eoscsecretariat.eu/working-groups/fair-working-group.

[15] S.J. Coles, J.G. Frey, E.L. Willighagen & S.J. Chalk. Taking FAIR on the ChIN. Data Intelligence 2(2020), 131–138. doi: 10.1162/dint_a_00035.

[16] Webpage of the Rare Disease Implementation Network. Available at: https://www.go-fair.org/implementation-networks/overview/rare-diseases/.

[17] Refinements of the Matrix questions: Available at: https://osf.io/xqfb9/.

[18] Dutch Techcentre for Life Sciences & Czech Technical University in Prague (2016). Data Stewardship Wizard (DSW). Available at: https://ds-wizard.org/.

[19] S. Jones, R. Pergl, R. Hooft, T. Miksa, R. Samors, J. Ungvari, R.I. Davis & T. Lee. Data management planning: How requirements and solutions are beginning to converge. Data Intelligence 2(2020), 208–219. doi: 10.1162/dint_a_00043.

[20] The ENVRI-FAIR project website. Available at: http://envri.eu/envri-fair/.

[21] GO FAIR Reference Ontology. Available at: https://github.com/go-fair-ins/GO-FAIR-Ontology.

[22] Semantic data models of FAIR Implementation Choices and Challenges. Available at: https://w3id.org/fair/icc/terms/FAIR-ICC-Model.

[23] G. Guizzardi & J. Mylopoulos. Taking It to the Next Level: Nicola Guarino, Formal Ontology and Conceptual Modeling. Frontiers in Artificial Intelligence and Applications 316(2019), 223–241. doi:10.3233/978-1-61499-955-3-223.

[24] B. Mons, H. van Haagen, C. Chichester, P. Hoen, J.T. den Dunnen, G. van Ommen ... & E. Schultes. The value of data. Nature Genetics 43(2011), 281–283. doi:10.1038/ng0411-281.

[25] T. Kuhn. Broadening the scope of nanopublications. In: P. Cimiano, O. Corcho, & V. Presutti et al. (eds) The Semantic Web: Semantics and Big Data. ESWC 2013. Lecture Notes in Computer Science, vol 7882. Berlin: Springer, 2013. doi: 10.1007/978-3-642-38288-8_33.

[26] T. Kuhn, A. Meroño-Peñuela, A. Malic, J.H. Poelen, A.H. Hurlbert, E.C. Ortiz ... & M. Dumontier. (2018). Nanopublications: A growing resource of provenance-centric scientific linked data. In: 2018 IEEE 14th International Conference on e-Science (e-Science). doi:10.1109/eScience.2018.00024.

[27] FAIR vocabulary. Available at: https://w3id.org/fair/principles/terms/FAIR-Vocabulary.

[28] T. Kuhn, E. Willighagen, C. Evelo, N.Q. Rosinach, C. Laura & I. Furlong (2017). Reliable granular references to changing linked data. In: d'Amato C. et al. (eds) The Semantic Web – ISWC 2017. ISWC 2017. Lecture Notes in Computer Science, vol 10587. Cham: Springer, 2017. doi:10.1007/978-3-319-68288-4_26.

[29] Global Soil Partnership. Information available at: http://www.fao.org/global-soil-partnership/en/.

[30] Encyclopaedia of Life. Available at: https://eol.org.

[31] Global Biodiversity Information Facility. Available at: https://www.gbif.org.

[32] International Chemical Identifier. Available at: https://www.inchi-trust.org.